

**A Comparison of Parent and Teacher Ratings
Of Children's Behaviors**

Michael W. Firmin

Elizabeth Proemmel

Chi-en Hwang

Cedarville University

Previous studies have compared the accuracy of parent, teacher, and clinician ratings of children behavior, especially in diagnostic analysis. However, many have questioned the validity of the tests and the value of each rater. While some research has found differences among raters, few had looked at samples of non-referred children. We wanted to study "normal" children, and we hypothesized finding no significant difference between the two raters. In our study, we administered the Clinical Assessment of Behavior to teachers and parents of students ranging from six to eighteen years old. When comparing these ratings, we found, as hypothesized, the parent and teacher ratings of children's behavior to possess statistically significant agreement. The only domain with significant disagreement was the externalizing domain. We found several potential causes for these findings consistent with previous research and suggested areas for further research, especially regarding the importance of the various raters for children who have not been clinically referred.

When children are screened for mental disorders, psychologists may use several methods, including clinical evaluations, interviews, and rating forms (Achenbach, 2001; Epkins, 1995; Powers, et al., 1998). According to Lengua, Sadowski, Friedrich, and Fisher (2001) and El-Hassan Al-Awad and Sonuga-Barke (2002), two widely-used rating forms to determine problem areas include Achenbach's *Child Behavior Checklist* (CBCL) and the *Conners' Rating Scales* (CRS). Both of these sets of instruments allow input from different informants (child, care-giver, and teacher) before making a diagnosis, and some researchers have found the agreement between raters to be within generally acceptable psychometric parameters (El-Hassan Al-Awad & Sonuga-Barke). Other researchers have found significant variability among different raters, as child-behavior ratings "can produce highly divergent results in individual cases" (Youngstrom, Loeber, & Stouthamer-Loeber, 2000, p 1046). Researchers and practitioners alike query the causes of this

In a theoretical sense, the variance in ratings may be related to the underlying classifications of disorders. In their research, Hartman, et al. (1999) concluded that there is no solid evidence to support syndromes as they are currently used in an operational construct. They noted, in particular, that similar questions on different rating scales can result in divergent diagnoses. Lengua, et al., (2001) also discussed problems with the dimensions on the CBCL, including the overlap of items and items that do not conceptually fit a dimension. Others have been unsatisfied with using a dichotomous diagnosis of either present or absent (e.g., Achenbach, 2001; Youngstrom, et al., 2000). This dichotomy controls processes in research and diagnosis, making variations in sources difficult to resolve. Achenbach criticized the syndromes as measuring different aspects of the child (mental health, personality, ability to adapt, etc.), instead of a more consistent categorization. Towers, et al. (2000) also recommended research addressing the distinction between answering questions based on the child's personality or on any occasional actions of the child. He emphasized the utilization of methods that are more empirical in order to identify those who may be on the borderline of a diagnosis, a caution also espoused by Drotar, Stein, and Perrin (1995). Along with the dichotomy of diagnoses, the scales on rating forms (Towers, et al.) are often only three-point scales (0 for "not true for child" to 2 for "very often true for the child") and tend to be skewed toward the better behavior, which can create a floor or ceiling effect (Hartman, et al.).

In a more practical sense, the variance in ratings may be related to the unique perspectives that the raters bring to the task. The question then becomes whose perspective is more accurate for diagnosis, which is challenging to determine because therapists seldom use teacher ratings (Towers et al., 2000). Schmitz, Saudino, Plomin, Fulker, and DeFries (1996) observed that, compared with parent raters, teachers have a standard-of-comparison advantage in evaluating children due to their regular interaction with a broad range of children. They can compare various children of the same age, bringing each child's behavior into a better perspective. Epkins (1995) and Wrobel and Lachar (1998) also highlighted the advantage of observing a particular child interacting with other children in both structured and unstructured settings, which can be especially important for a supported rating in the social skills dimension. Culp et al. (2001) also established the significant impact of the educational level of the teacher compared to the parent, where teachers are often asked to complete evaluations, but the parent's lack of education may make it difficult to complete.

The benefit of having groups of children can also be a drawback to accurate ratings, however, as teachers have less time for individual observation (Epkins, 1995). Towers, et al., (2000) recognized that, of the two broad categories often used for behavior, externalizing behaviors (behavior related to others) will receive more attention than internalizing behaviors (behaviors related to one's self) because of their disruptive nature. They also acknowledged that teachers' ratings would vary as a result of the demographic characteristics of the children being rated. Youngstrom et al. (2001) found that adolescents' race and socioeconomic status specifically predicted less agreement between teachers and parents on internalizing problems.

Even though parent evaluations of their children may be reasonably reliable for information, some researchers raise concerns about parents' subjectivity in the rating process (Schmitz et al., 1996). A main concern for parents' ratings is their own psychopathology, specifically depression (Towers et al., 2000). Youngstrom, et al. (2001) collected ratings from parents, teachers, and children, along with extensive demographic data. They found that parental depression caused the parents' ratings of both internalizing and the externalizing problem behavior to increase disproportionately with the same ratings by teachers and children, although the researchers also acknowledged that teachers' depression could have an effect on ratings.

Another concern with regard to parents' ratings is the increased tolerance of some behaviors (Loeber, et al., 1990). El-Hassan Al-Awad and Sonuga-Barke (2002) proposed that low reports of child problem behavior by the parents might be a result of more lenient standards of their own children or even stigma avoidance. Meydith, Prout, and Blaha (2003) also studied parents' tendency to respond with socially desirable answers, finding that parents may underreport maladaptive behavior, especially on the externalizing scales. Other causes for low reports of problem behavior may include comparison between one's own children, leading to inflated differences (Towers, et al., 2000), cultural differences (Youngstrom et al., 2000; Drotar et al., 1995) or simple ignorance (Loeber, Green, & Lahey, 1990) on the part of the parents.

Researchers also note that various contexts might confound low correlations between parent and teacher results. Wrobel and Lachar (1998) proposed that behavior problems vary across settings. The discrepancy causes concern about the relative validity of test scores deriving from any single source, because the difference in ratings may mask a significant degree of problematic behavior in children.

Conversely, Schmitz, et al. (1996) argued that dispositions may not actually change in different settings, but the settings provide varying views of the same disposition. Regardless of the cause, if behavior truly varies across settings, then multiple sources are essential to discerning specific problematic behaviors in the specific settings and treating them as needed (Achenbach, 2001; Youngstrom et al., 2000).

In summary of what is known to date, clinicians assessing children with behavioral disorders or who possess psychiatric disorders can expect the agreement between raters to vary based on the educational level of the parent, psychopathology of the parent or teacher, race and socioeconomic status of the child, and the settings available for observation. Progressing from these findings, we conducted a study which we believe helps to fill a gap in the research literature. In particular, a salient need exists relating to assessing children without behavioral problems and/or who have not been diagnosed with psychiatric disorders. The studies conducted to date have focused on parent/teacher rating forms for children with suspected pathology present in the children. There are numerous occasions when psychologists must make appraisals of non-diagnosed children using standard rating forms. Such instances include child custody evaluations, family assessments, foster care appraisals, adoptions, and the like.

Consequently, our present study focused on non-at-risk children being rated by teachers and parents who possessed no known psychiatric disorders. Our research project obtained data assessing the differences between teacher and parent ratings of such children. We believe this is important since sometimes psychologists must make evaluations based, in part, on data from rating forms gained from either teachers or parents, but not both. Our research question focused on how these two informants typically would differ from one another, when rating the same normal children. The applicability of the findings from this research also provide some expected norms by which psychologists using parent/teacher rating forms can expect parent and teacher ratings to differ, when practicing in non-psychiatric milieu and conditions.

Based on the data in the literature relating to children with behavioral problems or psychiatric disorders showing differences in parent/teacher ratings, we hypothesized finding no significant differences in our present sample. The present study was exploratory research, however, since nothing has been reported with this type of population to date.

Method**Participants**

We selected 33 students from a public school located in a small, rural, Midwest town. We chose the sample based on a cross-section of students in stratified populations of students in this particular school. One child chose not to participate. Also, teachers declined to participate in two of the ratings; therefore, those children were not included in the study. Forms with incomplete data (ten or more unanswered questions) were excluded, which resulted in an analysis of 28 participants. These 28 students ranged between six and eighteen years old, with a mean of age 11.08. In the three cases where the reported age for the child by the teacher was inconsistent with the reported age by the parent, we assumed that the parents' answer would be more accurate. Of the sample, 13 were female and 15 were male. All but two of the students lived in a rural area, and all but one of the students were Caucasian. This student was classified as Asian by the teacher and "other" by the parent.

Most of the parent raters were mothers, but two of the parent raters were the fathers. Their ages ranged from 32 to 47, with a mean age of 39.54. All of the parents were Caucasian and had at least 12 years of schooling, with the median number of years being 16. The teachers were also all Caucasian and mostly females, with six of the raters being male. Their ages ranged from 22 to 57, with a mean age of 42.79. All of the teachers had at least 16 years of education, with the median being 19. A majority (61%) of the teachers lived in a rural area, 29% lived in a suburban area, and 10% lived in an urban area.

Materials

We used the *Clinical Assessment of Behavior* (CAB), published by Psychological Assessment Resources (Braken & Keith, 2000). The parent form includes 260 questions in six domains: externalizing behaviors (59), internalizing behaviors (46), social skills (60), competence (47), adaptive behaviors (19), and critical items (29). The teacher form includes 125 questions in four domains: externalizing behaviors (40), internalizing behaviors (17), social skills (36), and competence (32). The questions were answered by completing a five-point Likert scale. The raters filled-out A for always or very frequently, B for often, C for occasionally, D for rarely, and E for never. For the scoring process, most questions were related to negative behaviors with 1 for never, 2 for rarely, 3 for occasionally, 4 for often, and 5 for always or very frequently. For the questions regarding positive behaviors, the scale was reversed. Therefore, a lower score represents a better behaved child.

Procedure

We obtained parental consent from all of the children who participated in this study. The surveys were hand-delivered directly to the parents at their homes, the teachers in their offices, and picked-up within 48 hours. The retest was distributed and picked-up two weeks later in the same manner.

Results

First, we calculated the mean ratings of parents and teachers for each of the categories. If a child's score was not available on a particular item, we imputed the personal mean scale score by taking the total raw score and dividing it by the total number of questions answered from that scale. As seen in Table 1, the mean for the parents' ratings ranged from 1.05 to 1.91, and the mean for the teachers' ratings ranged from 1.36 to 1.77. The standard deviations for the parents' ratings ranged from .06 to .37, and the standard deviations for the teachers' ratings ranged from .33 to .49.

Table 1: Mean and Standard Deviation of CAB Scores for Parent and Teacher Raters

<i>Behavior categories</i>	<u>Parent raters</u>		<u>Teacher raters</u>		<i>t(27)</i>
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	
Competence (test)	1.77	0.29	1.71	0.49	0.81
Externalizing behaviors (test)	1.64	0.37	1.40	0.34	3.43**
Internalizing behaviors (test)	1.88	0.37	1.74	0.39	1.66
Social skills (test)	1.91	0.32	1.77	0.38	1.87
Adaptive behavior (test)	1.61	0.33	----	----	----
Critical items (test)	1.06	0.09	----	----	----
Competence (re-test)	1.71	0.29	1.67	0.41	0.62
Externalizing behaviors (re-test)	1.57	0.32	1.36	0.33	4.16***
Internalizing behaviors (re-test)	1.80	0.33	1.71	0.40	1.03
Social skills (re-test)	1.82	0.35	1.75	0.39	1.23
Adaptive behavior (re-test)	1.57	0.34	----	----	----
Critical items (re-test)	1.05	0.06	----	----	----

** p<.01, *** p<.001

For the paired sample tests comparing the parent with teacher ratings, the t-value for externalizing behavior ratings was 3.43 (p<.01) on the test and 4.16 (p<.001) on the retest. The paired sample t-values

for social skills, competency, and internalizing scores showed a directional relationship, but did not reach the level of statistical significance.

Parent and teacher ratings on the test showed a correlation of .56 ($p<.01$) for the competence domain. The externalizing behavior correlation was .46 ($p<.05$). The internalizing behavior correlation was .31. The social skills correlation was .46 ($p<.05$). All correlations were statistically significant, except internalizing behavior. This pattern was also true for the retest.

Table 2: Correlations between Categories of Child Behavior as rated by Teachers

Category	1	2	3	4
1. Competence	0.89	0.66	0.59	0.83
2. Externalizing	0.60	0.82	0.45	0.80
3. Internalizing	0.55	0.45	0.80	0.64
4. Social Skills	0.71	0.56	0.55	0.85

Note: Test-retest reliability for teachers is in bold font. The correlations for each category are above the diagonal for the test, and below the diagonal for the retest.

Table 3: Correlations between Categories of Child Behavior as rated by Parents

Category	1	2	3	4	5	6
1. Competence	0.88	0.65	0.59	0.64	0.54	0.29
2. Externalizing	0.71	0.92	0.76	0.90	0.33	0.63
3. Internalizing	0.65	0.77	0.80	0.73	0.41	0.78
4. Social Skills	0.76	0.85	0.78	0.81	0.31	0.65
5. Adaptive	0.52	0.38	0.40	0.44	0.92	0.25
6. Critical		0.330	.580	.680	.530	.10 0.74

Note: Test-retest reliability for parents is in bold font. The correlations for each category are above the diagonal for the test, and below the diagonal for the retest.

We also calculated Pearson correlations for each of the categories against the other five for both the tests and retests (see Tables 2 and 3 for a summary). The strongest correlation for the parents occurred between social skills and externalizing behavior on the test and on the retest. The strongest correlation for the teachers occurred between social skills and competence on the test and on the retest.

Discussion

As we hypothesized, the overall ratings on the children's behavior for both parents and teachers were similar. The only statistically significant difference found in the paired sample test existed on the externalizing behavior domain. We believe that this discrepancy may be due, at least in part, to our small sample size and the small range of responses. The discrepancy on ratings for externalizing behavior likely would be minimized in a larger and more diverse sample. In our sample of relatively well-behaved children parents may expect their children to be very well behaved and monitor their actions closely, while teachers may not even notice smaller misbehaviors if the actions do not cause any disturbance in a classroom full of kids.

We also found the correlations between parent and teacher ratings to be statistically significant. Internalizing behavior, which has no external criteria available, was the only domain where this was not the case. The lack of consistency may be due to the relatively short time that the teachers had known the children, as Culp et al. (2001) established as a potential moderating variable. Because teachers completed their ratings in the fall, children had only been in class for two or three months. However, the high correlations between the test and retest indicate that, knowing they would be expected to complete the rating forms again, the teachers could have looked at the children more specifically but still did not change their ratings to a significant degree.

In exploring the tests' reliability and validity, the test-retest correlations were high. However, the lack of divergent evidence between the externalizing behavior and social skills as rated by the parents brings the validity of the parent test into question on those two measures. Divergent evidence is defined as a low correlation between two measures that should not be specifically related, as defined by Merydith, Prout, and Blaha (2003). Nevertheless, because the teachers' ratings did not have this same high correlation, the problem may lie in the fact that parents do not often see their children interacting with other children, and, as Epkins (1995) argued, teachers have more opportunities to observe children in social settings.

Culp et al. (2001) stressed that the perspectives of different raters are essential to a successful treatment plan. Different raters may be observing different behaviors, especially based on the systematic differences they found. The different perspectives makes each rater important (Valk et al., 2001), and psychologists should look at each

rater individually instead of just the ratings in aggregate form (Drotar, 1995).

Limitations and Need for Future Study

The present study is limited on a number of levels. Our sample size was relatively small, which has obvious implications for both external validity and statistical power considerations. Moreover, we assessed children, parents, and teachers with one appraisal instrument. Conducting this research with multiple assessment inventories would enhance our study significantly and provide triangulation for the findings. Replication is required before we have full confidence in our findings also. This is necessary prior to practitioners being able to reasonably suppose that parent and teacher rating would be interchangeable.

We recommend further research on children who are at behavioral risks, to see if the opposite result occurs. When children show behavioral problems, do they appear to show this maladaptive pattern equally for both the parents and teachers? That is, do parents or teachers expect less from these children? Which group will notice more negative behavior? The answers to these questions could further refine the relationship between parent and teacher ratings on child behavior that we report in the present study.

References

- Achenbach, T. M. (2001). Challenges and benefits of assessment, diagnosis, and taxonomy for clinical practice and research. *Australian & New Zealand Journal of Psychiatry*, 35, 263-271.
- Braken, B. A. & Keith, L. K. (2001). *Clinical Assessment of Behavior*. Florida: Psychological Assessment Resources, Inc.
- Culp, R. E., Howell, C. S., Culp, A. M., & Blankemeyer, M. (2001). Maltreated children's emotional and behavioral problems: Do teachers and parents see the same things? *Journal of Child & Family Studies*, 10, 39-50.
- Drotar, D., Stein, R. E. K., & Perrin, E. C. (1995). Methodological issues in using the Child Behavior Checklist and its related instruments in clinical child psychology research. *Journal of Clinical Child Psychology*, 24, 184-192.
- El-Hassan Al-Awad, A. M., & Sonuga-Barke, E. J. S. (2002). The application of the Conners' Rating Scales to a Sudanese sample: An analysis of parents' and teachers' ratings of childhood

- behaviour problems. *Psychology & Psychotherapy: Theory, Research & Practice*, 75, 177-188.
- Epkins, C. C. (1995). Teachers' ratings of inpatient children's depression, anxiety, and aggression: A preliminary comparison between inpatient-facility and community-based teachers' ratings and their correspondence with children's self-reports. *Journal of Clinical Child Psychology*, 24, 63-70.
- Hartman, C. A., Hox, J., Auerbach, J., Erol, N., Fonseca, A. C., & Mellenbergh, G. J. et al. (1999). Syndrome dimensions of the Child Behavior Checklist and the Teacher Report Form: A critical empirical evaluation. *Journal of Child Psychology & Psychiatry*, 40, 1095-1116.
- Lengua, L. J., Sadowski, C. A., Friedrich, W. N., & Fisher, J. (2001). Rationally and empirically derived dimensions of children's symptomatology: Expert ratings and confirmatory factor analyses of the CBCL. *Journal of Consulting & Clinical Psychology*, 69, 683-698.
- Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology*, 19, 136-143.
- Merydith, S. P., Prout, H. T., & Blaha, J. (2003). Social desirability and behavioral rating scales: An exploratory study with the Child Behavior Checklist/4-18. *Psychology in the Schools*, 40, 225-235.
- Power, T. J., Andrews, T. J., Eiraldi, R. B., Doherty, B. J., Ikeda, M. J., & DuPaul, G. J. et al. (1998). Evaluating attention deficit hyperactivity disorder using multiple informants: The incremental utility of combining teacher with parent reports. *Psychological Assessment*, 10, 250-260.
- Schmitz, S., Saudino, K. J., Plomin, R., Fulkner, D. W., & DeFries, J. C. (1996). Genetic and environmental influences on temperament in middle childhood: Analyses of teacher and tester ratings. *Child Development*, 67, 409-422.
- Towers, H., Spotts, E., Neiderhiser, J. M., Hetherington, E. M., Plomin, R., & Reiss, D. (2000). Genetic and environmental influences on teacher ratings of the Child Behavior Checklist. *International Journal of Behavioral Development*, 24, 373-381.

- van der Valk, J. C. van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (2001). Using parental ratings to study the etiology of 3-year-old twins' problem behaviors: Different views or rater bias? *Journal of Child Psychology & Psychiatry*, 42, 921-931.
- Wrobel, N. H., & Lachar, D. (1998). Validity of self- and parent-report scales in screening students for behavioral and emotional problems in elementary school. *Psychology in the Schools*, 35, 17-27.
- Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting & Clinical Psychology*, 68, 1038-1050.